DOCUMENT RESUME

ED 464 412 EA 031 649

| | |
|---|---|
| AUTHOR | Gong, Brian |
| TITLE | Designing School Accountability Systems: Towards a Framework and Process. |
| INSTITUTION | Council of Chief State School Officers, Washington, DC. |
| SPONS AGENCY | Department of Education, Washington, DC. Planning and Evaluation Service. |
| PUB DATE | 2002-01-00 |
| NOTE | 37p.; A publication of the State Collaborative on Assessment and Student Standards (SCASS) Accountability Systems and Reporting (ASR) Consortium. |
| AVAILABLE FROM | Council of Chief State School Officers, Attn: Publications, One Massachusetts Ave., N.W., Suite 700, Washington, DC 20001 ($10). Tel: 202-336-7016; Fax: 202-408-8072; Web site: http://www.ccsso.org/publications. |
| PUB TYPE | Guides - Non-Classroom (055) |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | *Accountability; Elementary Secondary Education; Program Design; Program Evaluation; Program Improvement; State Departments of Education; State Programs |

ABSTRACT

This document presents three different views of accountability to address state needs as their departments of education design, improve, or review their state accountability and reporting systems. The first of three sections presents the system-design decision process as a linear sequence of ten steps from defining the purposes of the accountability system to defining what will be reported and how data will be combined to make an accountability judgment. The second section provides alignment questions to help a state consider the internal consistency of its existing accountability policies. The third section provides descriptions of states that exemplify major models of accountability designs as outlined in the previous two sections. An appendix provides a list of references, along with example documents that illustrate some tangible deliverables a state department of education might have to produce when designing and implementing an accountability system, such as federal laws affecting accountability, advisory committee recommendation reports regarding accountability and reporting system design, accountability systems implemented in law, and accountability system technical manuals. A roster of fall 2001 Accounting Systems and Reporting participants concludes the document. (Contains 24 references.) (RT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 412

EA

# Designing School Accountability Systems

## *TOWARDS A FRAMEWORK AND PROCESS*

THE STATE COLLABORATIVE ON ASSESSMENT AND STUDENT STANDARDS (SCASS)
ACCOUNTABILITY SYSTEMS AND REPORTING (ASR) CONSORTIUM

Brian Gong and ASR SCASS

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B. Butterbaugh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

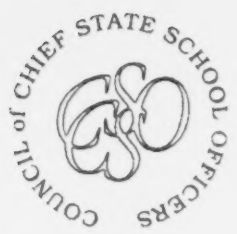
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS



EA031649




# COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nationwide, nonprofit organization composed of the public officials who head departments of elementary and secondary education in states, the District of Columbia, the Department of Defense Education Activity, and five extra-state jurisdictions. CCSSO seeks its members' consensus on major education issues and expresses their views to civic and professional organizations, to federal agencies, to Congress, and to the public. Through its structure of standing committees and special task forces, the Council responds to a broad range of concerns about education and provides leadership on major education issues.

Because the Council represents each state's chief education administrator, it has access to the educational and governmental establishment in each state and to the national influence that accompanies this unique position. CCSSO forms coalitions with many other education organizations and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

This report is sponsored by the CCSSO State Collaborative on Assessment and Student Standards (SCASS), Accountability Systems and Reporting (ASR) Consortium. Preparation of this report was supported in part by the U.S. Department of Education, Planning and Evaluation Service. The views and opinions expressed in this report are not necessarily those of the Council of Chief State School Officers, the U.S. Department of Education, or the State Collaborative on Assessment and Student Standards.

SUELLEN K. REED (INDIANA), PRESIDENT
MICHAEL E. WARD (NORTH CAROLINA) PRESIDENT-ELECT
PETER MCWALTERS (RHODE ISLAND) VICE PRESIDENT

G. THOMAS HOULIHAN
EXECUTIVE DIRECTOR

COPIES OF THIS REPORT MAY BE ORDERED FOR $10.00 FROM:

COUNCIL OF CHIEF STATE SCHOOL OFFICERS
ATTN: PUBLICATIONS
ONE MASSACHUSETTS AVE., N.W., SUITE 700
WASHINGTON, DC 20001
PHONE: (202) 336-7016
FAX: (202) 408-8072
OR GO TO WWW.CCSSO.ORG/PUBLICATIONS

Accountability Systems and Reporting

State Collaborative on Assessment and Student Standards

# DESIGNING SCHOOL ACCOUNTABILITY SYSTEMS:

## *Towards a framework and process*


BRIAN GONG
Center for Assessment
*and*
ASR SCASS


January 2002


Rolf K.Blank
Jennifer G. Manise
ASR SCASS Coordinators



Funding to support development of this paper has come from ASR member states and the U.S. Department of Education, Planning and Evaluation Service.


© Council of Chief State School Officers


Washington, DC


75 Years of Excellence in Education

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

# CONTENTS


PREFACE 1

The Accountability Systems and Reporting SCASS ..... 1
Objective: Assist States With Accountability Design ..... 1
Organizations: Three Views of State Accountability Systems ..... 2

SECTION 1: COMPREHENSIVE QUESTIONS ABOUT ACCOUNTABILITY DESIGN 3

Overview of the Design Decision Process ..... 3
Empirical and Policy Analyses ..... 3
Description of Design Decisions ..... 4
1. What Are the Purposes of the Accountability System? ..... 4
2. What Are the Main Contexts, Political and Otherwise? ..... 7
3. What Are the Main Legal and Policy Constraints? ..... 9
4. What Are the Units of Measurement, Accountability, and Reporting? ..... 9
5. What Are Schools/Students (or Others) to be Held Accountable For? ..... 11
6. What Accountability Decisions Will Be Made, and With What Consequences? ..... 13
7. How Will Results Be Reported? ..... 14
8. What Data Are Available and Will Be Used in the Accountability System? ..... 15
9. How Will Data Be Combined to Make an Accountability Judgment? ..... 17
10. How Will the Accountability System Be Monitored and Evaluated? ..... 18

SECTION 2: ALIGNMENT OF MAIN ACCOUNTABILITY ELEMENTS 19

Introduction ..... 19
How to Use ..... 19
Limitations ..... 19
Alignment Question 1: For what will schools be held accountable? ..... 20
Alignment Question 2: Does the state have sufficient and appropriate data? ..... 21
Alignment Question 3: What other state policy requirements are important? ..... 22

SECTION 3: EXAMPLES OF REPRESENTATIVE APPROACHES TO ACCOUNTABILITY DESIGN 23

Examining Models ..... 23
Texas ..... 23
Kentucky, Louisiana, and California ..... 24
North Carolina ..... 24
25

APPENDIX: ANNOTATED EXAMPLES AND REFERENCES FOR RESOURCES FOR ACCOUNTABILITY DESIGN 27

Selected Resources ..... 27
Recommendations By Advisory Committees Regarding the Design of Accountability Systems ..... 27
Accountability Systems Implemented In Law ..... 27
Manuals ..... 27
References, Citations, and Resources ..... 28

ROSTER OF ASR PARTICIPANTS *AS OF FALL 2001* 31

# PREFACE

## *The Accountability Systems and Reporting SCASS*

The ASR (Accountability Systems and Reporting) SCASS (State Collaborative on Assessment and Student Standards) is one of several collaborative projects initiated by the Council of Chief State School Officers (CCSSO). The ASR SCASS project is working to develop documents that will help state departments of education design, improve, or review their state accountability and reporting systems. Increasingly, state departments of education respond to differing needs for data and reports that serve bottom-line accountability requirements and provide useful information to educators, policymakers, and the public. The ASR-member state departments of education began working together in 2000 to improve the departments knowledge of designs for accountability and reporting methods across the states and the effectiveness of the systems based on differing designs. The first priority established by the ASR group was to develop a document that would assist states in making decisions about designs for accountability systems.

### OBJECTIVE: ASSIST STATES WITH ACCOUNTABILITY DESIGN

The ASR SCASS representatives[1] began by developing a framework for considering the range of issues and topics that most states have to address. Priority topics for consideration in accountability design identified by the ASR SCASS state members included the following:

- definition of good school and associated accountability models;
- survey of outcome measures used for accountability and means to deal with multiple measures ;
- consequences (especially rewards and sanctions);
- reporting;
- assistance models; and
- evaluation of the impact of accountability.

As this document was being finalized, Congress passed the No Child Left Behind Act of 2001 legislation reauthorizing the Elementary and Secondary Schools Act (ESEA). This legislation provided more extensive federal requirements for states student assessment and school accountability systems than had previously existed. The legislation stipulated assessments in reading and mathematics (and eventually science) in grades 3-8 and specified that states must develop an accountability system with at least certain assistance and sanction provisions. These requirements were extended to all schools, not only schools receiving Title I assistance, as had previously been the case. As is usual, states are now expecting the U.S. Department of Education to issue rules or guidance to clarify several specific aspects of how the law should be implemented and how it might apply to the states individual circumstances. It is beyond the scope of this paper to consider the new ESEA legislation and possible rule and/or guidance interpretations.

The states participating in the ASR SCASS have a wide range of experience and are at different stages of designing and implementing state accountability systems. Some have

[1] See Appendix for list of ASR-member states and representatives.

already implemented detailed systems and are involved in dealing with the results and fine tuning their systems. Some states have a good idea of how they would like to approach the design of their accountability systems but have not yet worked out the details or committed to the design in regulation, statute, or operational programs. Some states are in the early stages of identifying their constraints and reviewing their options for designing an accountability system. The states requested different types of information, organized in different ways, to meet their varied needs.

## Organizations: Three Views of State Accountability Systems

This document presents three different views of accountability design to address states needs. One view presents an elaborated framework, with questions, criteria, and comments, intended to provide a structure for helping states move through the process of designing a school accountability system The second view presents a concise checklist of characteristics to help states evaluate the consistency and coherence of existing programs. The third view provides examples of actual state experience with design features that might be considered and why.

**Accountability Design Decisions**. The first section of the paper presents the design decision process as a linear sequence of ten steps from defining the purposes of the accountability system to defining what will be reported and how data will be combined to make an accountability judgment. Each design step is discussed in some detail, which is especially useful for states with little experience in school accountability design or for policymakers seeking a more comprehensive understanding. States may not always follow all of these steps in this order, but the list is intended to be comprehensive so that states can see where they fit and identify their needs. (We focus on schools; we do not address the design of student or district accountability systems, although many of the same topics may be relevant.)

**Coherence of Policies**. The second section provides alignment questions to help a state consider the internal consistency of its existing accountability policies. This section focuses on key decisions regarding what a school should be accountable for, available data, inclusion, and reporting. This view is especially useful to states, which are considering modifying their accountability systems or are reviewing the consistency of their systems.

**State Examples**. The third section of this document provides descriptions of states that exemplify major models of accountability designs as outlined in the previous two sections. These real world references are useful in understanding operational details, relationships, rationales, and contexts for evolving and implementing policy.

References to important resources for accountability design appear throughout this document. The **Appendix** provides a complete list of the citations, along with example documents that illustrate some tangible deliverables a state department of education might have to produce when designing and implementing an accountability system, such as:

- federal laws affecting accountability;
- recommendation reports by advisory committees regarding the design of accountability and reporting systems;
- accountability systems implemented in law, either as a set of statutes passed by the state legislature or as a set of regulations passed by a state board of education; and
- accountability system technical manuals.

# SECTION 1

# COMPREHENSIVE QUESTIONS ABOUT ACCOUNTABILITY DESIGN

## Overview of the Design Decision Process

The process outlined in this section represents a step-by-step, logical approach to designing an accountability system:

1. What are the purposes of the accountability system?
2. What are the main contexts, political and otherwise?
3. What are the main legal and policy constraints or specifications?
4. What are the units of performance, accountability, and reporting?
5. What are schools/students (or others) to be held accountable for?
6. What accountability decisions will be made, and with what consequences?
7. How will results be reported?
8. What data are available and will be used in the accountability system?
9. How will data be combined to make an accountability judgment?
10. How will the accountability system be monitored and evaluated?

The design decision process presented as a linear sequence of ten steps will be especially useful for states with little experience in school accountability design or for policymakers seeking a more comprehensive understanding. Design decisions, however, are usually complex, with many interacting assumptions and relationships. A state would likely follow a more iterative and perhaps less restrictive, step-by-state process than the sequence portrayed here.

### Empirical and Policy Analyses

The design process should be checked with empirical analyses and reviewed with policymakers to ensure that the evolving design can be implemented acceptably. For example, states should perform reliability analyses to ascertain that the level of error or uncertainty associated with accountability decisions is acceptable to the DOE and to key policymakers. Relatively few states have conducted such studies, and those who have often do not make them public. However, it is clear that states need this type of information for legal and professional defensibility of high-stakes programs. The Appendix provides document sources and web addresses of related criteria, practical standards, and sample studies. It is highly recommended that states thinking of conducting empirical studies contact a state department that has already established a program of research and evaluation.

## Description of Design Decisions

### 1. What Are the Purposes of the Accountability System?

**General Purposes**

Often state accountability systems have general purposes, such as

- *to identify and promote improved educational practices and results;*
- *to inform stakeholders of the condition of education at the school, district, and state levels and to identify areas in which improvement is needed and success is being achieved;*
- *to obtain the support of all stakeholders in making the changes needed to enable all students to achieve at high levels; and/or*
- *to inform policy decisions and actions by officials at the local, state, and federal levels, parents, students, members of the community, and other interested individuals to improve academic performance where needed and to reward it where appropriate.*

**Specific Purposes**

Accountability systems report school performance on variables or indicators, as do the report cards issued by many states. School accountability systems differ from report cards in that they

- focus on the school as the unit of reporting, whereas many state report cards use the state as the unit of analysis;
- focus on student performance, whereas many state report cards report a wide range of input and descriptive variables; and, most importantly,
- report school performance in relation to criteria or standards established by the state, thereby providing a legal and credible operational system for evaluating and publicizing school performance results and assigning rewards, assistance, and sanctions.

By reporting performance in relation to standards, school accountability systems are intended to identify good- and low-performing schools. It is important to note that there are at least four main conceptual definitions of good- or high-performing. [2] It is essential that the state clearly identifies which specific purpose(s) or definition of good it intends its accountability system to reflect.

---

[2] The formulation of these four dimensions follows work done by Dale Carlson and Richard Hill (Personal communications, April-October 2001). Previous presentations by Carlson, Hill, and Gong noted the differences between status, improvement, and growth, which correspond to the top two cells and bottom row of what is presented in this document. Hill (2001) and Gong (2001) have investigated the technical characteristics, especially reliability, of school accountability systems representing the four cells. See also papers by Bob Linn, which additionally describe adjustments for SES: *Reporting school quality in standards-based systems* (2001); and *Accountability models* (2001), CRESST paper presented at ECS annual meeting.

**Table 1: Criteria for Defining Quality School in an Accountability Model**

| | Criteria of "Good" | |
|---|---|---|
| | **Status** | **Change** |
| **Achievement** *(in relation to standards)* | **Model 1:** How high do students in the school score on state assessments? What percentage of students meets the state standards? | **Model 2:** Is the school improving, or increasing, the performance of classes of students over time? Is the percentage of students meeting the state standards increasing from one year to the next? |
| **Effectiveness** *(in relation to past performance of students)* | **Model 3:** Are students learning as they progress through the grades? Are individual students making expected progress from grade to grade? | **Model 4:** Is the school becoming *more* effective—is it helping students (individuals, subgroups, or all) reach higher levels of achievement or learn relatively more over the years than was achieved or expected in the past? |

Another way to express the above definitions of quality is to apply the following models to the stem, In the accountability system, a good school is one where . . .

- a high percentage of students meets the standards (*Model 1:status of achievement*).

  For example, a Commended School might have 70% of its students meet or exceed the state standard for proficiency, and a Low School might have 50% of its students meet or exceed the standard.

- the percentage of students meeting the standard is increasing (*Model 2: change of achievement*).

  For example, a Commended School might have 40% of its students meet or exceed the state standard for proficiency in year 1, and 50% of its students meet or exceed the standard in year 2; and a Low School might have 60% of its students meet or exceed in year 1, but 50% meet or exceed in year 2.

- a high percentage of students make progress during the year, in relation to where they started, regardless of whether or not the students meet the standard. (Model 3: This is called effectiveness since it relates to how well the school does with the student inputs it receives.).

  For example, a Commended School might have students score at the 2.5 grade level at the end of grade 3, and the same students score at the 3.5 level at the end of grade 4 (i.e., the students made one grade level growth from the end of grade 3 to the end of grade 4); and a Low School might have students score at 3.2 in grade 3 and 4.0 in grade 4 (i.e., students grew, but less than the expected one grade level amount).

- the progress made by students during one year, in relation to where they started, is higher than the progress made by students the previous year (*Model 4: change in effectiveness in other words, the school is becoming more effective over time*).

  For example, a Commended School might have students score at the 2.5 grade level at the end of grade 3 in year 1 and at the 3.5 level at the end of grade 4 in year 2 (growth of 1.0), *and* have scores of 2.2 at the end of grade 3 in year 2 and 3.3 at the end of grade 4 in year 3 (growth of 1.1). A Low School would have less growth between years 2 and 3 than it had between years 1 and 2.

Note: No state defines a good school according to Model 4. Model 4 is included in this discussion primarily for conceptual completeness. Reasons for the model s unpopularity include the arguments that change (improvement) is not linear and that schools should be held accountable for changes in the rate of change. States may not have adopted the model because such systems appear more complex. In addition, Model 4 is less reliable than the other models, primarily because the amount of change to be detected is relatively small.

The design of the accountability system particularly, what data are collected, how they are combined, and how they are interpreted is *critically* linked to the definition (or purpose) chosen as the focus or emphasis of the accountability system. The validity of the accountability system will depend upon this stated purpose or definition of quality and how well the system reflects this purpose.

**Comparison Groups**

There are, of course, many variations to these four basic definitions of quality in Table 1. One important variation that could be applied to any of the four models is a comparable group requirement. Simply stated, the accountability system may include a requirement that the performance of the school be comparable to some other group. Two common comparison groups are discussed below:

A. *Subgroup comparison (e.g., racial/ethnic subgroup).* A typical requirement would be that all subgroups perform comparably to the school as a whole. This approach has the effect of requiring the school to meet the same (or nearly the same) standards for all subgroups. The main reason for requiring comparable growth for subgroups is to ensure that schools are accountable for equitable results, so that disparities between subgroups are not hidden by aggregated averages. No Child Left Behind requires these comparisons.

Although it could conceivably be applied to all four models, the subgroup comparison has been applied prominently in Models 1 (status achievement, as in Texas) and 2 (status improvement, as in California). A drawback to this approach is that it usually makes the accountability system much less reliable in a statistical sense, because subgroups involve fewer students than does the school as a whole, and fewer students lead to less reliable accountability decisions.[3]

B. *Comparable schools comparison.* In this approach, schools are grouped together based on prior achievement of students and/or common demographic characteristics of the students/schools. A school s performance is then compared to the other schools in the group rather than to an absolute standard. States have typically created comparable school groups based on a combination of characteristics, often including some indicator of SES, race/ethnicity, mobility, and other factors that usually correlate with achievement.

The main reason for using comparison school groups is to provide a context for interpreting results. This approach can also enable schools to seek help from higher-performing schools with similar demographic characteristics. A drawback to using a school comparison is that it usually is incompatible with an approach of common standards for all students. That is, the comparable schools approach usually means that poor, non-white students are expected to score lower than students in schools with less challenging demographic backgrounds. Research shows that using comparable schools may be more appropriate for Model 1 systems (status achievement) and less appropriate

[3] See Hill (2000), *The reliability of California's API*, for an empirical analysis of the reliability of one state's requirement that subgroups make improvements comparable to the school as a whole.

for Model 2 and 3 systems, since improvement of schools or growth of students may be less correlated with SES or race/ethnicity.[4]

## 2. What Are the Main Contexts, Political and Otherwise?

Educational changes, such as instituting an accountability system, take place within a complex context. The following questions illustrate important political, legal, cultural, and other contextual circumstances that the design of an accountability system should consider.

*What are the main reasons the state has come to the point of considering an accountability system?* For example, is accountability . . .

- linked to a financial equity lawsuit (e.g., we re providing more funding so we better make sure we re getting our money s worth and equity in student outcomes is a known parameter by which the system will be evaluated and reviewed )?
- a drive to increase what students can do (e.g., we want high school diplomas to mean something; we want to better prepare students for the modern world of work; we want to eliminate social promotion )?
- a preface to taking strong action (e.g., let s give one more chance to schools demonstrating really poor quality, and then a sanction such as reconstitution, consolidation, charters, or vouchers should be considered )?
- a means for addressing inequities between schools or subpopulation groups (e.g., we need to ensure that schools do a better job of educating traditionally underserved groups, schools that have been historically disadvantaged in the state, etc. )?
- a way to validate a generally strong educational system and challenge it to improve its capacities?

*What are the existing legal requirements in statute or regulation?* For example:

- State legislation/court orders
  - Are there specific aspects mandated by legislative or judicial institutions that the assessment system or accountability system must include or address? (If so, these aspects must be included in the accountability system design, and they usually are more difficult to change.)
- Federal legislation (ESEA/Title 1, IDEA97, etc.) (States must comply with federal laws, although many states have gotten waivers for specific requirements in the past.)
- State education regulations

*What are the cultural norms of the state?*

- How urgently is change expected? How much time is reasonable to see results?
- How much change is expected? How much improvement is perceived as needed?
- How broadly is change expected does the accountability system apply to K-8, high school? Is it set within a context of P-16 reform?
- How centralized is the state is there a tradition of high definition of curriculum and assessment by the state?

[4] See, for example, early results from Kentucky reported in the Kentucky *KIRIS Technical Manual*. See Appendix for full reference.

- What is the relationship between the state department of education and local education agencies? What is the role of the state, districts, school boards, schools, students, the public, the business community, the legislature, and other stakeholders?
- How committed is the state to inclusion of all students in the accountability system and to common standards of performance for all schools? New ESEA requirements under No Child Left Behind requires that all students be included and standards apply to all schools and students.
- Does the state department of education have a clear ethical stance of what it considers right and good in terms of educational outcomes and means to achieve those ends?
- How susceptible is the accountability system to change by political pressure?

*How technically inclined is the state?*

- How sophisticated or complex an accountability system is acceptable?
- How much technical capacity does the state department of education have?
  - Will the state contract out most work or do much of the data processing in-house?
  - Does state department have staff with sufficient time and statistical expertise to check contractors work and/or explore alternate accountability designs?

*How much money is the state willing to devote to accountability, and over how much time*

- for implementing an assessment system?
- for implementing an accountability system (especially providing assistance, rewards, and other consequences associated with accountability)?
- for infrastructure (e.g., student data bases, reporting, and staffing)?

*How much capacity does the state have in . . .*

- political will and leadership to sustain accountability/reform efforts?
- support among state board of education members, school superintendents, school administrators, teachers, business and community leaders, parents, professional associations, and other special interest groups, etc.?
- state department of education?
- contractors?

*Are there conflicts for example, among mandates, between mandates and purpose of system, between mandates and capacities? What mechanisms exist to resolve conflicts and solve problems?*

Note that the political and educational context of each state will be unique. What is possible at any point in time will likely differ from state to state. States should therefore be cautious about copying another state s system without first determining how well that system fits its own context. A strategic plan for an accountability system should also consider creating and maintaining the conditions (e.g., political, legal, operational capacity) necessary for a sound system.[5]

[5] See Consortium for Policy Research in Education (2000) for a profile of each state's accountability system, including state context. CCSSO is now reporting shorter annual profiles of state accountability systems together with state indicators of performance and context (see Manise et al. 2001).

## 3. What Are the Main Legal and Policy Constraints?

*Has the state legislature or state board of education established specifications for the accountability system? Specifications include*

- an accountability system will be established, by a certain date;
- the accountability system will incorporate certain data (e.g., student scores on state assessments);
- schools will be given specific ratings, designations, or labels (i.e., specifies what the possible labels will be); and
- certain consequences will depend upon the accountability system (e.g., assignment of sanctions or rewards).

*Has the state determined how extensive an accountability system it will establish or promote?*

- Will there be student and/or district accountability programs in addition to school accountability?

The interaction of student and school accountability can be quite complex. There is a core philosophical debate, as well, as to whether schools should be help-accountable if students have no stakes to do their best, and whether students should be faced with sanctions before schools have been held accountable to adequately prepare them. There are technical issues as well, such as whether an assessment is valid for both student decisions and school decisions. This document focuses on school accountability design but acknowledges that student and district accountability are essential topics to consider, especially since a growing number of states have initiated student accountability systems, particularly for high school.

*Has the state decided how the state system will meet federal law?*[6]

- Under new ESEA requirements Title I assessment provisions will be extended to all schools, and there must be one system of accountability for all public schools.
- How will students who participate in an alternate assessment be included in the school accountability system?

## 4. What Are the Units of Measurement, Accountability, and Reporting?

A state must decide on the unit the accountability system will focus on in measurement of performance, reporting of results, and accountability consequences. The units of measurement do not have to be the same as the units of accountability consequences. For example, it is common to have schools held accountable for the performance of specific grades of students (e.g., grades 3, 6, 8), and how students in those grades perform determines the consequences for the whole school. It is important that the units of measurement, accountability, and reporting be coherent.

The unit of measurement of performance represents the levels of data aggregation and

[6] As this paper was being finalized, the U.S. Congress passed the reauthorization of the *Elementary and Secondary Education Act*, including provisions for Title I. The new ESEA legislation has considerable changes for assessment and accountability systems. Those newly enacted provisions will need some elaboration and clarification by rules yet to be issued by the U.S. Department of Education.

disaggregation. Often decisions related to the unit of measurement reflect both considerations of purpose (validity) and operational concerns, such as how much testing time is acceptable, how many tests the state can afford, and what data are already available (e.g., attendance may be reported to the state at the school, not student, level).

The unit of accountability represents the persons or organizations the state holds responsible for performance within the accountability system. School accountability differs fundamentally from systems of district, teacher, or student accountability, because the attribution of results and assignation of consequences are focused on the school as an organization rather than on individuals or the district. It is absolutely essential that a state come to agreement about the unit of accountability, or the system will be seen as unfair, unjust, and unsupportable.

The unit of reporting represents how accountability results and performance data will be summarized and disseminated. Aggregation and disaggregation of results usually are intended to inform interpretations and actions. For example, the state may provide student-level data, not because it is holding students accountable, but because it facilitates the school in analyzing its curriculum, instruction, and student support patterns.

**Table 2: Common Units of Measured Performance, Accountability Consequences, and Reporting for School Accountability Systems**

| *School Accountability System* | *Unit(s) of Measurement* | *Unit(s) of Accountability Consequences* | *Unit(s) of Reporting* |
|---|---|---|---|
| state | | | P |
| region | | | P |
| school-level (e.g., elementary, middle, high) | | | P |
| individual content area | | | P |
| demographic subgroup | | | P |
| district | | | P |
| administrators in district | | ✓ | P, S |
| school | | ✓✓✓ | P, S |
| all teachers in school | | ✓✓ | P, S |
| school principal | | ✓ | P, S |
| all school administrators | | ✓✓ | P, S |
| groups of teachers (e.g., grade 4 teachers, algebra teachers) | ✓ | ✓ | P, S |
| individual teachers | ✓ | • | S |
| demographic subgroups of students in school | ✓ | ✓ | P, S |
| grades (classes) of students | ✓✓✓ | | S |
| individual students | ✓✓✓ | • | S |
| content areas/ standards | ✓✓ | | P, S |
| subtest scores | | | S |
| item scores | | | S |
| comparison groups | ✓✓ | | P, S |
| time span (history or trend) | ✓✓✓ | | P, S |

Frequency across states: ✓ = occasional ✓✓ = common ✓✓✓ = very common
P = reported publicly S = reported to school personnel; usually not publicly known

* This paper focuses on school accountability. Several states have indicated their intention to also implement student accountability systems, with consequences for individual students in terms of promotion, graduation, or diploma endorsement. Several states have teacher accountability systems.

Table 2 lists common units, or levels, of measurement, accountability consequences, and reporting. The table indicates relative frequency of use across states currently.[7] The most important point to note is that there is little overlap between the three areas, particularly between units of measurement and units of consequences. This underscores the need for the state to clearly conceptualize and communicate the accountability system to ensure agreement with the inferences being made between performance and accountability (e.g., why it is fair to hold teachers accountable for students performance). As also shown in Table 2, the school usually has more information to analyze and act on than has been publicly released.

On a more pragmatic note, the accountability system must have clear definitions of each unit. For example, states commonly have to reconcile different school definitions that have been established for funding, administrative, and accountability purposes. Similarly, a common issue for the state to define is when a school is accountable for a student whether a student shows up on the day of testing (or moves part way through testing to another school), has been enrolled in the school for the full year, or something in between.

Accountability definitions have tremendous implications for the design of the program in terms of what data are gathered and how they are reported. This accountability question is discussed in further detail below.

## 5. What Are Schools/Students (or Others) to be Held Accountable For?

In Step 1, we considered the purposes of the system and criteria for defining a quality school. Now, we go further in setting standards for school performance.

*In the accountability system, which standard determines a good school?*

1. A high percentage of students meets the standards (status of achievement).
2. The percentage of students meeting the standards is increasing (change of achievement).
3. A high percentage of students scores higher at the end of the year than where the students scored the previous year, regardless of whether or not the students meet the state proficiency standard (status of improvement over time).
4. The percentage of students making progress during the year, in relationship to where they started, increases; or the amount of progress made increases over the previous year (change in improvement over time).

*Conversely, in the accountability system, which standard determines a bad school?*

1. A low percentage of students meets the standards (status of achievement).
2. The percentage of students meeting the standards is decreasing, or not improving quickly enough (change of achievement).
3. A low percentage of students makes progress during the year, in relationship to where the students started, regardless of whether or not the students meet the state proficiency standard (status of improvement over time).

---

[7] CCSSO has published survey summaries of the various accountability and indicator reporting across the states. See, for example, CCSSO (2000). *State Education Accountability Reports and Indicator Reports: Status of Reports Across the States*, which includes 50-state information on units of reporting. For a more detailed treatment, see Jaeger and Tucker (1998). *Analyzing, disaggregating, reporting, and interpreting students' achievement test results: A guide to practice for Title I and beyond.*

4. The percentage of students making progress during the year, in relationship to where they started, decreases; or the amount of progress made decreases over the previous year (change in improvement over time).

One of the most important tasks in implementing an accountability system is determining which model or purpose the state believes in. The state must then set standards or criteria for what is acceptable. For example, let us say a state has determined that it will define school performance quality in terms of the percentage of students who meet or exceed the state standards each year on the state tests (status of achievement). The state must then define what is passing, and what percentage of students passing constitutes a high performance. States have been challenged to set these accountability criteria in ways that are rigorously demanding, yet educationally realistic and politically acceptable. States have chosen different ways to do this. Texas, for example, started with a low requirement of 50% of students passing and increased the percentage to 80% over a number of years. Kentucky set the standard very high, created intermediate goals, and gave schools 20 years to meet the long-term goal.

It is possible for a system to combine models of quality a system may incorporate multiple definitions of good and may not have strictly parallel definitions of good and bad. For example, Louisiana, Kentucky, and California are examples of states with basic accountability systems that focus on school improvement over time (status change). In a pure status-change system, every school would be expected to improve, and every school that improved would get some credit. In fact, it is common to establish an upper bar of achievement, such that a school that had high-performing students would not be expected to improve, and a lower bar, such that a school with very low-performing students would be identified, regardless of how much it had improved.

Accountability systems have had to meet other requirements. A key source of guidance (and requirements) for accountability systems has been the federal Title I program (see footnote 6). Since 1994, Title I has required states to institute assessment and accountability systems for schools served by Title I. Over the past several years, most states have tried to unify their systems for Title I with state assessment and accountability. As that has happened, Title I has had a large influence in moving states toward systems that:

- have provisions to include all students in the assessment and accountability systems;
- incorporate multiple measures, including assessments of higher-order thinking skills, often interpreted to mean involving test formats other than or in addition to multiple-choice;
- use standards-based performance levels to describe student performance;
- establish performance standards for schools involving all students moving toward or meeting standards of proficiency ;
- disaggregate accountability results at school and district levels by student groups including race/ethnicity, LEP, SES, disability, and migrant;
- establish a state definition of adequate yearly progress ;
- require states to identify and support schools in need of improvement; and/or
- require establishment of a district as well as a school accountability system.

Note that the first three requirements concern the nature of the assessments. Much effort at the state and federal levels in the past four to five years has been expended trying to

define and meet these requirements and to evaluate states efforts to do so. Major efforts have included developing alternate assessments for special education students, expanding state assessments to include constructed-response items, and defining standards-based content and performance frameworks.

The latter five requirements involve accountability systems. Until very recently, these requirements have received less federal guidance (other than to require states to develop something), and they have been subject to much wider variation in interpretation among states. The federal legislation for the reauthorization of the *Elementary and Secondary Education Act* (passed at the time this document was finalized in December 2001) creates more extensive specifications regarding these areas, including minimal content areas, frequency of assessment, adequate yearly progress, and accountability consequences for low-performing schools. It is anticipated that the U.S. Department of Education will issue further rules to offer guidance in much more specificity. This will be especially important for states with existing assessment and accountability systems.

It is useful to note that for many people, a good school for accountability purposes is not necessarily the opposite of a bad school, as defined by the accountability system. For example, many people agree that a school is bad (or low-performing) if the majority of its students cannot reach a minimum standard in reading and math. However, many people also agree that a good school does more than teach its students to read and do math. The implication is that an acceptable accountability system may need to pay close attention to defining quality not only in terms of what is valued, but also in terms of how it is expressed at the ends of the continuum representing high/good and low/bad. In other words, a system that is adept at identifying low-quality schools may not necessarily identify high-quality schools in a way that agrees with people s experience or values.

## 6. What Accountability Decisions Will Be Made, and With What Consequences?

Every current state accountability system involves reporting a public designation, label, or rating. Indeed, making an evaluative judgment in relation to some standard is what distinguishes current accountability systems from school report cards and other descriptive systems. Such descriptive systems, available for years, have published a wide variety of data, but they have not assessed performance in terms of what is good enough and have not attached consequences to performance.

In addition to *describing* and *evaluating* schools, a state will need to decide whether there will be other *consequences*.

These accountability decisions will need to be made by the state:

- identification of and assignment of labels to high- and low-performing schools (e.g., distinguished schools, schools in need of improvement);
- assistance and/or sanctions to schools in need of improvement (e.g., additional funds, targeted professional development, school support teams, requirement to follow a school improvement plan, corrective actions, student transfer, faculty evaluation, reconstitution);
- rewards to high-performing schools (e.g., funds, waivers from regulations, identification to provide technical assistance, citations or other public recognition).

The state should have a sound rationale for making these decisions and should put forth a rationale for what educational consequences it expects as a result of the accountability

decisions. The validity of the accountability system will be evaluated, in part, upon the consequences of the accountability decisions.

It is especially important that the state describe the specific uses for the accountability information it is reporting, and how different users (e.g., parents, teachers, administrators, policymakers) might apply the accountability information toward improvement. For example, a state may expect that schools will improve sufficiently through local mechanisms spurred only by public reporting. With that expectation for accountability, the state should outline the scenarios for those local mechanisms, such as:

- Student achievement, and school accountability scores, will improve through strong curriculum alignment with the state standards (supported by the state establishing high-quality standards sufficiently specified that schools can align instruction to the standards)
- Achievement and accountability scores will improve through public pressure on local schools by parent and community involvement (e.g., specify who gets the data, how to pressure schools to serve all students equitably and adequately).
- Achievement and accountability scores will improve through the threat of parents requesting their children be transferred between teachers, public school buildings, or to charter schools or other alternatives.

When the state establishes consequences within its accountability system rewards, support systems and assistance, or sanctions it is even more important that it construct a rationale for what impact those consequences are expected to have, and how.

At the time this document was being finalized there was little available research or history on the impact of various rewards, assistance, or sanctions programs used by states. This was in part due to the fact that few states had more than a few years of experience in assigning consequences, and in part due to the limited number of systematic studies done. Several organizations had announced intentions to study the impact of specific consequences (as contrasted with the implementation of an assessment or an announced accountability system). The Consortium for Policy Research in Education (CPRE), for example, has published reports on the implementation and effect of the monetary rewards system in Kentucky during its first five years (Kelley, 1999).

## 7. How Will Results Be Reported?

Results are reported to inform understanding and action. What is reported should be linked to the view of who will take action and how (see preceding steps 4 and 6).

A state will need to decide whether to report many different possible indicators, such as:

- single overall rating or label;
- multiple ratings or labels (e.g., status, improvement);
- relation to other schools (e.g., comparison bands);
- numeric accountability score(s) (e.g., status score, improvement target, improvement score, overall score, score on component parts such as each content area test);
- results for subgroups (e.g., rating, accountability score(s), assessment score(s));
- information on inclusion and participation;
- previous accountability and/or assessment results (e.g., historical or trend data);
- elements that are reported but are not included in determining accountability results.

In addition, the state will need to decide what it should report for which unit of analysis (e.g., student, teacher, grade, school, district, state) and for what time span.

With increasing use of the web and other software tools, states have more options in terms of deciding how much detail to report, in what form, and with what interpretive support.[8] It is clear that the trend for reporting is both toward making more raw information available and making more syntheses and tools available to help people make sense of the data.

## 8. What Data Are Available and Will Be Used in the Accountability System?

In a conceptual design, after the purposes, uses, and other questions are answered, the question of specific data should be addressed. Of course, practical concerns usually make this an iterative discussion at best. Questions to consider include:

- What data are available or could be available?
- What data will be incorporated into the accountability system to determine accountability results?
- What data will be reported but not used for accountability? Do the accountability results depend on any calculations or interim results not reported publicly?
- What factors will influence the inclusion of data into the accountability system?

  Examples of criteria[9] to be considered when selecting data are discussed below.

- Suitability of data for accountability purposes
  - Are pre-/post-test scores needed? Do scores need to be tracked to individual students over time? Do scores need to be tracked to individual teachers? (See Section 2.)
- Validity of measures, including alignment of assessments with state content and student performance standards
  - States considering using commercial off-the-shelf tests should ascertain whether the tests are adequately aligned with state standards to provide valid measures and to influence instructional alignment in a constructive manner.
  - More states are giving systematic attention not only to content specifications (how specific? how extensive? how public?), but also to the skills and cognitive complexity required by items. The discussion about constructed response and performance assessments is shifting from face validity to more principled analyses. States should consider whether their frameworks include adequate specificity in terms of content *and* performance standards,

[8] The ASR SCASS is currently working on the issue of reporting and should have some helpful documents available in the future. CCSSO has available a *Professional Development for Assessment Literacy* CD-ROM that addresses uses of assessment data in reporting, which have considerable overlap with issues of accountability reporting. ECS has a project on "second generation accountability models" that deals extensively with innovative reporting mechanisms. See Appendix for references.

[9] Standards for accountability systems are still evolving. This list of criteria for data expands on the set of criteria (validity, fairness, credibility, utility) developed by Eva Baker, CRESST, *Standards for accountability systems*, available at www.nciea.org. A similar presentation, *Watching the watchers: Standards for accountability systems* (Baker & Linn, 1999), is available at the CRESST website, www.cse.ucla.edu/CRESST/conf99/bakeroh/sld001.htm

and what types of assessments are needed to provide valid information.

- Reliability of measures and results
- Understandability, usefulness, and credibility of results
- Frequency and scope of data collection (e.g., annually, grade levels)
- Timing of data collection (e.g., spring, summer, fall) in relation to accountability reporting and usage
- Cost of data development, processing, and reporting

States typically have used student test scores as performance indicators in accountability systems. Some states have included non-test indicators[10] as well.

Factors for consideration in the discussion of performance indicators include:

- Types of statewide student assessments and content areas covered
  - States consider many factors when deciding whether to test more subjects than reading and math. Of increasing relevance is whether schools narrow their curriculum to match the tested areas and thus inappropriately reduce instruction in subjects such as science, social studies, arts, music, or physical education.
  - Several states are moving toward end-of-course tests and away from survey tests, particularly in high school. End-of-course tests require strong specification by the state of content to be taught by grade level and course; many local control states do not have (nor wish to exert) such curricular influence. End-of-course tests and survey/census tests each have logistical demands, as well as their own set of accountability issues.
- Assessments to include all students (e.g., accommodations, students with limited English proficiency, special education students/students with disabilities)
  - Federal regulations require appropriate assessments be provided for all students. Most states have complied with developing an alternate assessment for students with moderate to severe disabilities who cannot participate in the regular assessment with all accommodations and whose IEPs/504 plans prescribe an alternate assessment. However, many states face challenges of deciding upon and providing appropriate assessments for other subgroups, including students with limited English proficiency and other students who currently must take the regular assessments with modifications that invalidate their results.
  - Federal law, under ESEA (HR.1), is clear that all students should be included in the assessment and reporting of assessment results: Until now some states have exclude large groups of students through various assessment and/or accountability policies. For example, some states excluded from accountability (although not from assessment) students who had not been in the district or school for at least one year. Some states assessed students but allowed modifications that invalidated the assessment results and excluded such results from accountability.

[10] CCSSO, ECS, and CPRE have good summaries of what indicators states have included in their accountability and reporting systems. See the Appendix for references. See also Erpenbach, Carlson, LaMarca, and Winter (2001).

- Other student performance indicators (e.g., dropout/persistence rates; graduation rates; student attendance; teacher attendance; percentage of teachers with certification in assigned field; class size; students per teacher in secondary schools; measures of school climate; safe schools; parental and community involvement)
- Most states currently have or have had a report card with results of indicators other than test scores. States must decide whether to include both rated and non-rated elements on a single report card or whether to issue multiple reports.
- Non-test indicators should have certain technical qualities, such as commonality across schools, suitable variance and reliability, validity, and availability within the desired time schedule.

## 9. How Will Data Be Combined to Make an Accountability Judgment?

Despite the magnitude of the data brought into the system, every state s current accountability system boils those data down into just one accountability decision or judgment (or two, if status and improvement are reported separately, and there is no overall label). As a result, every state s current school accountability system includes multiple pieces of data that must be combined to make an accountability judgment. Common types of data combination include:

- student scores on the same assessment within a grade and content area (e.g., grade 4 math on the state assessment) to produce a grade/school score or rating;
- student/school scores across grade/content areas to produce a school score or rating (e.g., grades 4 and 8 math, grades 4 and 8 reading);
- student/school scores across years (e.g., average or difference of grade 4 in one year with grade 4 in previous year);
- test scores/ratings with other assessment and/or non-test scores (e.g., test scores and portfolios, attendance and dropout);
- scores/ratings for status and improvement (e.g., overall score equals status plus two times improvement);
- past and current scores/ratings analysis to determine current accountability status (e.g., give a more severe rating or consequences to a school identified as low-performing two years in a row).

An important reason for combining scores is to increase the stability and reliability of the decisions made on the basis of the scores. A second reason is to simplify the system for accountability decision-making and reporting.

In making accountability judgments, data that are unlike can be combined as well. Prominent approaches used by states to combine unlike data (e.g., different content areas, assessment instruments, or groups of students) include:

- an indexing system that assigns points and combines them into an overall score (e.g., $x$ points for status plus $y$ points for improvement = $z$ points overall );
- a rule-based system that describes how combinations map to accountability judgments and consequences (e.g., If a school has at least $x$% of its students meeting or exceeding the standard, *and* if the subgroups in the school made significant improvement, *then* the school shall be designated a successful school );

- a formula with weights (e.g., a multiple regression formula such as overall score $= a + weight_b variable_1 + weight_c variable_2 . weight_d variable_3$).

The accountability system can also combine multiple dimensions[11] (e.g., status and improvement) in the following ways:

- multiple ratings/dimensions may be combined into a single overall rating;
- multiple ratings may be reported, but each combination is associated with a single accountability consequence;
- multiple ratings may be given, with multiple consequences possible.

If the accountability system seeks to contextualize school performance by considering other factors (e.g., prior achievement, demographics), then additional data must be combined. Taking into account prior achievement or demographic variables of students usually involves a statistical approach, such as multiple regression.

## 10. How Will the Accountability System Be Monitored and Evaluated?

A state should create and follow a plan to monitor and evaluate its accountability system. Some key concerns are identified here.

- Is the system complete?
- Can the system be improved?
- Is the system having the desired effects?
- Is the system producing undesired effects?
- Have assumptions or circumstances changed to an extent that the system should change?

The *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 1999) provides useful guidance, especially for assessment systems. The *Program Evaluation Standards* (JCSEE, 1994) is another strong source of guidance for relevant criteria for evaluation in general. Suzanne Lane (1999) provides a good overview with practical examples of validity studies actually conducted for a state assessment program. The Kentucky accountability system technical manual for 1999[12] also offers some good examples of analyses to monitor an accountability system in its early stages.

Unfortunately, few states have committed appropriate resources and energy to evaluation, which would help maintain accountability system credibility and utility. This is an important area for all states to commit appropriate resources and attention.

---

[11] The CCSSO CAS SCASS is working on papers that address multiple measures and the issues of how to combine scores to produce accountability judgments. See especially Erpenbach et al. (2001). Gong (2001) discusses the tension exerted by validity's call for more extensive samples of broad domains and reliability's need for focused, repeated administrations of the same instrument. See Appendix for full citations.

[12] The most recent edition may be obtained from the Division of Assessment and Accountability, Kentucky Department of Education, 500 Mero St., Frankfort, KY.

BEST COPY AVAILABLE

# SECTION 2

# ALIGNMENT OF MAIN ACCOUNTABILITY ELEMENTS

## Introduction

The purpose of this section is to help a state reflect on the alignment of its (proposed) accountability design to ensure that the system is internally consistent.

The alignment should include three main areas:

- definition of what schools will be accountable for;
- data requirements; and
- other policy requirements, particularly inclusion and reporting.

## How to Use

To check for alignment, the state should follow the steps below.

1. Beginning with alignment question 1 in the following table, the state should specify for what schools should be held accountable. This standard should align with one of the three models described.
2. For each part of alignment questions 2 and 3 in the following tables, the state should identify its situation. Staying within columns indicates greater internal consistency or alignment. Movement across columns, on the other hand, indicates some mixture and less alignment of models, purposes, or capacities.

## Limitations

As noted in the text, it is possible to combine these models. Section 3 provides some examples of states that have used variants or combinations.

To be comprehensive, these tables would include additional topics, notably accountability consequences, suggested technical analyses, and more detailed discussion of how to deal with the myriad issues of implementing an operational system. Such detail is beyond the scope of this paper, which is intended to provide a starting point for states.

## *Alignment Question 1: For what will schools be held accountable?*

| Alignment Question | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| What are schools accountable for? | How high do students in the school score on state assessments? What percentage of students meets the state standards? | Is the school improving, or increasing, the performance of classes of students over time? Is the percentage of students meeting the state standards increasing from one year to the next? | Are students learning as they progress through the grades? Are individual students making expected progress from grade to grade? |
| Hypothetical example of *Commended* rating: | School has 80% or more students meet or exceed proficiency standard. | School makes at least sufficient improvement to meet expected growth goal, e.g., school went from 20% of grade 4 students meeting proficiency standard in Year 1 to 30% of grade 4 students in Year 2, or from Index score of 55.0 in baseline year to Index score of 60.1 in growth year. | School had students on average make at least sufficient growth for the year, e.g., students made "one year's expected growth" between grade 4 and grade 5 (between year 1 and year 2). |
| Hypothetical example of *Low-performing* rating: | School has 50% or fewer students meet or exceed proficiency standard. | School did not make sufficient improvement to meet expected growth goal. | School's students did not make sufficient growth for the year, e.g., students made less than "one year's expected growth." |
| Variations | • Increase required standard over time, e.g., 50% in year 1, 55% in year 3, 60% in year 5.<br>• Require comparable performance of subgroups. | • Require minimal or no improvement for high-scoring schools; identify very low-scoring schools regardless of improvement.<br>• Require reduction in proportion of lowest scoring students.<br>• Require comparable improvement for subgroups. | • Determine expected growth by historically empirical growth versus by goal of where state wants to be.<br>• Require comparable growth by subgroups. |

## *Alignment Question 2: Does the state have sufficient and appropriate data?*

| Alignment Question | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| *Does the state have sufficient data?* | Does the state conduct annual testing in at least one grade level per school with at least Pass/Fail performance levels, and include all students? | Does the state have at least two years of data (baseline, growth) for at least one grade level per school with at least three performance levels, and include all students? | Does the state have at least two years of data (pre/post) for at least two successive grades per school, preferably with linked or comparable scales across years, and individual student tracking over years, and provisions to monitor inclusion? |
| *Number of grades of data* | At least one grade per school | At least one grade per school | At least two grades per school |
| *Number of years of data* | One | At least two years of data (baseline, growth); many states use four years | At least two years of data (pre/post) |
| *Grade placement* | No restriction | No restriction | Must be adjacent, e.g., grades 4 and 5 |
| *Types of tests* | Can mix assessments and content areas over grades (e.g., CRT in grade 4, NRT in grade 5, local assessments in grades 4 and 5; or CRT math in grade 4 and CRT reading in grade 5) | Can mix assessments and content areas over grades (e.g., CRT in grade 4, NRT in grade 5, local assessments in grades 4 and 5; or CRT math in grade 4 and CRT reading in grade 5) | Must have consistent content areas and preferably consistent assessment instruments every grade-pair |
| *Performance standards* | Minimum one cutpoint, e.g., Passing/Not Passing | At least three performance levels preferable (for reliability reasons) | Vertical or grade-linked scale scores preferable |
| *Student ID tracking* | Not necessary | Not necessary | Matching student pre- and post-test scores preferable, although could use quasi-longitudinal groups (e.g., scores from all students in grade 3 in year 1 and scores from all students in grade 4 in year 2) |
| *Data other than test scores* | One year; need minimum of Pass/Fail performance standard for each indicator | At least two years; need definition of desired improvement | Include non-test data using Model 1 or Model 2 approach |

## *Alignment Question 3: What other state policy requirements are important?*

| Alignment Question | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| ***Inclusion*** | Can include all students | Can include all students | Often does not include all students |
| *Special education students taking alternate assessments* | Can include if Pass/No Pass performance standard is set | Can include if performance standards and growth targets are set | Can include if comparable scales (often difficult) or growth targets are set, and if alternate assessment is administered every grade |
| *Mobile students* | Can include | Can include | Only students with both a pre- and a post-test score (unless using a quasi-longitudinal model comparing non-matched successive groups) |
| *LEP students* | Can include if Pass/No Pass performance standard is set, either on regular or non-English test | Can include if performance standard is set, either on regular or non-English test | Can include if comparable scales (often difficult) or growth targets are set |
| ***Reporting*** | | | |
| *Relation to standards* | Simple, direct relation to proficiency performance standard (Pass/No Pass) | Direct relation to student and school performance standards; relative growth target more complicated to understand | "Expected growth standard" more difficult to understand; may not be related to customary student performance proficiency standards |
| *Simplicity: single outcome* | Yes | Possible, although often states report status and growth components separately | Possible |
| *Decision frequency* | Annual | Can be annual; often biennial if two years of data are combined | Annual |
| *System start-up* | Requires one year | Requires at least two years of data, often four; state may implement provisional system until full data are available | Requires at least two years of data |
| *Can mix with other models* | Can mix with Model 2 and Model 3 | Can mix with Model 1; usually not combined with Model 3 | Can mix with Model 1; usually not combined with Model 2 |
| *Validity relation to SES* | Highly correlated with SES | May have low correlations with SES | May have low correlations with SES |
| *Relative reliability* | Typically high | Often moderate to low | Often moderate to low |

# SECTION 3

## EXAMPLES OF REPRESENTATIVE APPROACHES TO ACCOUNTABILITY DESIGN

### Examining Models

| Focus of Accountability System | | |
|---|---|---|
| | **Status** | **Change** |
| **Achievement** | **Model 1:** How high do students in the school score on state assessments? What percentage of students meets the state standards? | **Model 2:** Is the school improving, or increasing, the performance of classes of students over time? Is the percentage of students meeting the state standards increasing from one year to the next? |
| **Effectiveness** | **Model 3:** Are students in the school learning (scoring higher) as they progress through the grades? Are individual students making expected progress from grade to grade? | **Model 4:** Is the school becoming more effective—is it helping students (or subgroups) achieve more over the years than the same students achieved or were expected to achieve in the past? |

| | | |
|---|---|---|
| ***1*** | **Status (Model 1):** *How are current students in the school performing in relation to the standard?* For example, is there a high percentage of students meeting or exceeding the state student proficiency standard? | North Carolina, Texas |
| ***2*** | **Improvement (Model 2):** *Is the school getting better at helping successive groups of students meet the standards?* For example, are grade 4 students scoring higher this year than did the grade 4 students two years ago? | California, Kentucky, Louisiana, Massachusetts, Oregon, Vermont |
| ***3*** | **Student Growth (Model 3):** *Are students learning from year to year?* For example, how much higher did students perform at the end of grade 4 this year than they did at the end of grade 3? | North Carolina, Tennessee |
| ***4*** | **Change of Effectiveness (Model 4):** *Students or subgroups make more than expected growth, or rate of improvement increases.* (implied in "closing the gap" between subgroups' absolute performance) | None known |

## TEXAS

Texas is an example of a state accountability system built on **Status (Model 1)**. Texas assesses students in grades 2-11 using a custom state CRT. Each student is designated a performance label. The percentage of students meeting or exceeding the proficiency standard (PAC, or percent above cut) is calculated for each school. The school is assigned an accountability label based on its PAC.

Texas requires that subgroups perform at comparable levels for school ratings. Thus, the school as a whole and each subgroup must reach the PAC for a particular rating. For example, a school that has 55% of its total students tested meet or exceed the proficiency standard would also need all of its subgroups (white, Hispanic, African-American, etc.) to have at least 50% proficiency or above for the school to receive an Acceptable rating.

The school labels and standards in 2000 were Exemplary (90% or more of students met or exceeded student proficiency standard), Recognized (80%), Acceptable (50%), and Low-Performing (less than 50%). The required PAC has been increased over the past several years from 50% in 1994 to 70% in 1999. In addition, the Texas Assessment of Academic Skills test was replaced in 2001 by a test intended to be more rigorous; thus the requirements for students to meet the proficiency standard (and thereby the requirement for schools to meet the PAC) should have been raised over time.

Texas does have a provision that schools could be rated Acceptable through having adequate improvement, which was defined as one-fifth the difference between where the school was and a target standard (as of 2000, the Acceptable level). (Fewer than five schools had a rating change due to the improvement clause in 1999.)

Note that although Texas assesses every student in every grade 2-11 in every subject annually, it does not have a student longitudinal growth model. Texas does track and match individual students but uses that data to exclude students from school accountability if a student did not attend in the same district the previous year, that student is not included in the accountability system.

Texas has several other school accountability provisions. For example, the state has a rewards provision that is based upon a school s relative ranking within a group of comparable schools.

For a full description of the Texas school accountability system, see the *Texas School Accountability Manual*. (See Appendix for annotated reference.)

## KENTUCKY, LOUISIANA, AND CALIFORNIA

Kentucky, Louisiana, and California are examples of state accountability systems built on **Improvement** of successive groups, where the school is expected to raise the achievement of cohorts over time, e.g., grade 4 in year 3 is expected to be higher than grade 4 was in year 1 **(Model 2)**. These states all generate expected improvement based on how far a school is from achieving the state goal. The amount a school actually improves is compared to the expected improvement, and an accountability label is assigned accordingly.

Kentucky assesses students using a custom state CRT, an NRT (5% of total nominal weight), and includes other indicators for school accountability. Louisiana uses a custom CRT, an NRT (30% of total weight), and other indicators. California uses an NRT (customized for the state). In each state, each student is designated a performance label (e.g., in Kentucky: Novice, Apprentice, Proficient, Distinguished). Each performance level is assigned a number of points. A school index is calculated as a weighted average of points. In Kentucky and Louisiana, two years of data are combined to create a baseline

(pre) and growth (post) index score; California uses one year of data for baseline and one year for growth.

Each state has set an overall state goal and a time period for achieving that goal. In Kentucky, the goal is 100 on the index, which is equivalent to all students, on average, meeting the state proficiency standard by 2013. Louisiana and California have variants, e.g., Louisiana has set a 10-year goal of all students, on average, rating Basic, and a 20-year goal of all students, on average, rating Proficient. California s improvement expectation is based upon a 5% reduction in the gap between baseline and long-term (approximately a 20-year) goal.

Each school has a growth target that represents the amount the school must improve over time (every two years in Kentucky and Louisiana; and one year in California), to meet the long-term goal by the target date. Variants: Louisiana and California recalculate the growth target every cycle, reflecting the school s actual status. Kentucky calculates the growth target once for the 20 years.

A school that exceeds the expected growth receives rewards. California and Kentucky have implemented financial reward programs, and Louisiana passed a statute in 2001 establishing financial rewards as part of its school accountability system. A school that is far from meeting its expected growth is declared Low-Performing.

California requires that subgroups make 80% of the expected growth of the total school in order for the school to receive rewards. Kentucky requires that a school reduce the proportion of students at the lowest achievement level by at least 10% in order for the school to receive rewards.

Note that California tests every student, every year, in every content area, and tracks whether a student attended the same district the previous year. However, California does *not* have a student growth model. California uses the matched data to exclude students from school accountability only students who attended the same district the previous year are included in school accountability. (This means, for example, that in districts with grades 7-12, the scores of all grade 7 students are excluded from school accountability.)

## North Carolina

North Carolina combines both **Status (Model 1)** and **Student Growth (Model 3)** in its school accountability system. Schools are assigned accountability labels on the basis of their status (PAC). North Carolina also assigns schools accountability labels based on whether the students have made one year s expected growth. For example, a student who enters grade 5 reading at grade level 3.5 (one-and-a-half grades below level) and exits grade 5 at grade level 4.5 would receive credit for making one grade level growth. Schools are accountable for helping their students make at least one grade level of growth each year.

North Carolina s expected growth is based on the achievement levels of grades at a point in time; that is, they are not linked to the proficiency standard and reflect what was rather than what is desired. North Carolina s formula for determining expected growth is more complex than most states , and includes provisions for regression to the mean and rate of growth. Regression to the mean is the statistical observation that, all things being equal, students or schools with extreme scores will tend to score closer to the mean upon retesting. The rate of growth is an adjustment that higher (or lower) performing schools historically could be expected to increase performance more (or less) than the average. In North Carolina s system, these opposing factors almost cancel each other out. The regression adjustment adds points to lower scoring schools, while the rate of growth factor gives fewer points; the opposite is true for higher scoring schools.

The student growth provision applies to students who attend the same schools within the same districts for at least two years (to provide pre- and post-test scores). The student growth portion of the accountability system thus excludes mobile students. Other assessment and accountability provisions also determine which students are included in accountability.

Note that schools are not responsible for closing the gap or bringing the student up to grade level in the student growth portion of the accountability system. The status portion does reflect whether all students in the accountability system are reaching proficiency.

# APPENDIX:

# *ANNOTATED EXAMPLES AND REFERENCES FOR RESOURCES FOR ACCOUNTABILITY DESIGN*

## *Selected Resources*

Much can be learned from the experience of states that have wrestled with designing and implementing school accountability systems. The documentation produced by these states can be invaluable sources of information. This Appendix provides examples of some key types of documentation design recommendations, statutes/regulations, and manuals that illustrate three different stages of implementation. This Appendix also provides references for additional resources.

### Recommendations By Advisory Committees Regarding the Design of Accountability Systems

Louisiana convened a commission to design its school accountability system. The DOE hired an advisory group to respond to the Commission s recommendations and provide guidance on how to implement the recommendations. This document is unusual because it provides a rare blend of policy rationale and technical and practical input. Available at: http://www.nciea.org/publications/LASchlDesign_TAC98.pdf

### Accountability Systems Implemented In Law

Accountability systems typically are implemented in some detail. Usually the system is formalized legally, either as statute or as regulations passed by the state board of education. This provides legal standing for enforcement. Statutes typically are more difficult to change, and the DOE typically has less close working relations with the legislature than it does with the state board. These examples show four different states approaches implementing their accountability systems in law.

1. Oklahoma (statute)
2. Kentucky (regulation) http://www.lrc.state.ky.us/kar/TITLE703.HTM
3. Louisiana (regulation)
4. Vermont (operations manual)

### Manuals

*Kentucky District Assessment Coordinator Guide/Accountability Manual* — provides a comprehensive source of detailed policies, procedures, and instructions for implementing the accountability system. Updated each year.

Available at: http://www.kde.state.ky.us/oaa/implement/DAC_Guide_2001-02/table_of_contents_2001.asp

*Texas Accountability Manual* — provides policies, procedures, and instructions for implementing the accountability system.

Available at: http://www.tea.state.tx.us/perfreport/account/2001/manual/

*Kentucky Technical Manual* — provides essential information about design, development,

implementation, validity, and reliability of the accountability system. Aimed at researchers, evaluators, and technical users.

## References, Citations, and Resources

This section contains the full reference for every resource cited in the paper. It also includes references for other good sources of information on accountability design.


American Education Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. The professional standards focused on assessment, less on accountability systems.

Baker, E. (2000). *Standards for accountability systems*. Presentation at the 2000 Reidy Interactive Lecture Series.

CRESST has been working on a set of standards for accountability systems, intended to be similar to the *Standards* for assessment. Available at www.nciea.org

Baker, E. & Linn, R. (1999). *Watching the watchers: Standards for accountability systems*.

Available at www.cse.ucla.edu/CRESST/conf99/bakeroh/sld001.htm

Center for Assessment (also known as NCIEA) is a non-profit company that has worked extensively on the design of school accountability systems. Staff have worked with several states on design and implementation of accountability systems. The Center also has produced a number of papers dealing with accountability design. A unique resource for practical information regarding accountability design decisions and guidance on implementation studies is the Reidy Interactive Lecture Series produced by the Center for Assessment (www.nciea.org), which includes criteria, procedures, and examples for critical aspects of accountability systems, including standard setting, validity studies for assessments, technical documentation, and reliability studies for accountability systems. The Center s website includes a centralized, easy access to all 50 state s departments of education, as well as to other professional organizations concerned with assessment and accountability in education.

Consortium for Policy Research in Education. (2000). *State assessment and accountability systems: 50 state profiles*. Philadelphia: University of Pennsylvania. Available at www.gse.upenn.edu/cpre/Publications/Publications_Accountability.htm

Council of Chief State School Officers. (2000). *State education accountability reports and indicator reports: Status of reports across the states*. Washington, DC: Author.

Includes 50-state information on units of reporting.

Available at www.ccsso.org/pdfs/AccountabilityReport2000.pdf

Council of Chief States School Officers, CAS SCASS. (2001). *Professional development for Assessment Literacy*. CD-ROM. Washington, DC: Author.

Erpenbach, W. J., Carlson, D., LaMarca, P. M., & Winter, P. (Eds.). (2001). *"Incorporating multiple measures of student performance into state accountability systems A compendium of resources.* "(Draft). Washington, DC: CCSSO.

Gong, B. (in press). *Multiple measures*. In S. Fuhrman & R. Elmore (Eds.). New York: Teachers College Press.

Gong, B., & Tappan, R. (2001). *How much school improvement should accountability systems require?*

This presentation, given at the 2001 Reidy Interactive Lecture Series sponsored by

the Center for Assessment, defines improvement for the four different accountability models discussed in this paper. The presentation presents results of relative reliability of the four models and looks at different data sources for trying to determine how much improvement is possible. It also presents case study results of dramatic, sustained improvements for some individual schools. Available at www.nciea.org

Hill, R. (2000). *The reliability of California s API.*

Reports an empirical analysis of the reliability (decision consistency) of California s school accountability system API (Academic Performance Index). Reports standard errors for different size schools and probabilities for decision consistency for different size schools, subgroups, and postulated amounts of true gain. One important finding is that for California s accountability system, using data available in 2000, the reliability of decisions based on subgroups was both low and biased against schools with diverse populations. Available at www.nciea.org

Hill, R. (2001). *Issues related to the reliability of school accountability scores.*

This paper, presented at the 2000 Reidy Interactive Lecture Series sponsored by the Center for Assessment, discusses issues of school reliability, including relationship between test score reliabilities and reliability of school accountability decisions, statistics for comparing accountability reliabilities, influence of calculation methods, and factors that affect accountability reliabilities. Intended for people familiar with terms used in assessment and reliability but who might not be familiar wita the issues of reliability when used in the context of accountability. Includes many tables for detailed examination and reference. Available at www.nciea.org

Jaeger, R., & Tucker, C. (1998). *Analyzing, disaggregating, reporting, and interpreting students achievement test results: A guide to practice for Title I and beyond.* Washington, DC: CCSSO. Available at www.ccsso.org pdfs analyze.pdf

The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs.* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Kelly, C., Odden, A., Milanowski, A., & Heneman, H. (2000, Feb.). *The motivational effects of school-based performance awards.* (Research Brief No. 29). Philadelphia: CPRE. Available at http://www.cpre.org/Publications/rb29.pdf

Kentucky Department of Education. (2000). *District assessment coordinator (DAC) manual for 2000-01.* Frankfort: Author.

Comprehensive source of implementation information about the accountability program. Available at
www.kde.state.ky.us oaa implement DAC_Guide_2000 table_of_contents_2000.asp

Lane, S. (1999). *Validity evidence for assessments.*

This presentation, given at the 1999 Reidy Interactive Lecture Series sponsored by the Center for Assessment, provides a framework for validity studies that should be conducted for state assessment programs. The framework is based on the 1999 *Standards for Educational and Psychological Testing.* Available at www.nciea.org

Linn, R. (2001). *Accountability models.* CRESST Paper, University of Colorado, Boulder. (Presented at ECS annual meeting.)

Linn, R. (2001). *Reporting school quality in standards-based systems.* (CRESST Policy Brief No. 3).
Available at http: cresst96.cse.ucla.edu CRESST Files policybriefnl.pdf

Manise, J., Blank, R., & Pewett, C. (2001). *State education indicators with a focus on Title I.* Washington, DC: U.S. Department of Education.

Texas Education Agency. (2001). *Texas school accountability manual.* Austin: Author.

Information by state on school accountability measures. A good example of a comprehensive source of information about a state school accountability program. The TEA website is an impressive example of many resources available online, including standards, policy, research, test objectives and items, technical documentation, and data.

Available at www.tea.state.tx.us/perfreport/account/2001/manual/

United States Congress. (1997). *Individuals with Disabilities Act* (IDEA97).

United States Congress. (2001). *Elementary and Secondary Schools Act, 2001.*

U.S. Department of Education. (1999, Nov.). *Peer reviewer guidance for evaluating evidence of final assessments under Title 1 of the Elementary and Secondary Education Act.* Washington, DC: Author.

# ROSTER OF ASR PARTICIPANTS

## Alaska

Mark Leal
Alaska Department of Education

Richard Smiley
Alaska Department of Education & Early Development

## California

Linda J. Carstens
California State Department of Education
Accountability

Patrick McCabe
California State Department of Education
Accountability

Bill Padia
California State Department of Education
Accountability

Cheryl Tiner
California State Department of Education
Accountability

Brian N. Uslan
California State Department of Education
Assessment

## Connecticut

Renee Savoie
Connecticut State Department of Education

Charlene Tucker
Connecticut State Department of Education

## Delaware

David Blowman
Delaware State Department of Education

Carole White
Delaware State Department of Education

## Louisiana

J.P. Beaudoin
Louisiana Department of Education

Bernadette Morris
Louisiana Department of Education

Scott Norton
Louisiana Department of Education

Margaret K. Singer
Louisiana Department of Education

## Minnesota

Cathy Wagner
Minnesota Department of Children, Families and Learning

## Nebraska

Bob Beecham
Nebraska Department of Education

Marilyn Peterson
Nebraska Department of Education

Pat Roschewski
Nebraska Department of Education

## Utah

Barbara Lawrence
Utah State Office of Education

Randy Raphael
Utah State Office of Education

Michael Taylor
Utah State Office of Education

## West Virginia

Jan Barth
West Virginia State Department of Education

Beth Cipoletti
West Virginia State Department of Education

Sandra McQuain
West Virginia State Department of Education

## Consultant

Brian Gong
Center for Assessment
PO Box 4084
Portsmouth, NH 03802-4084
603-766-7900 phone
603-766-7910 fax
bgong@nciea.org

## USDOE

Meredith Miller
US Department of Education
400 Maryland Avenue, SW
FB-6, Room 6W219
Washington, DC 20202
202-401-8368 phone
202-401-4353 fax
meredith.miller@ed.gov

Sue Rigney
US Department of Education
OESE/CEP
400 Maryland Avenue, SW
Room 3C139
Washington, DC 20202-6132
202-260-0931 phone
202-260-7764 fax
sue.rigney@ed.gov

Elois Scott
US Department of Education
Planning & Evaluation
400 Maryland Avenue, SW
Room 6W103
Washington, DC 20202
202-401-1958 phone
202-401-4353 fax
elois.scott@ed.gov

## CCSSO

Rolf Blank
Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington, DC 20001-1431
202-336-7044 phone
202-408-1938 fax
rolfb@ccsso.org

Jennifer Manise
Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington, DC 20001-1431
202-336-7029 phone
202-408-1938 fax
jenmm@ccsso.org

Wayne Martin
Council of Chief State School Officers
One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
202-336-7010 phone
202-789-1792 fax
waynem@ccsso.org

John Olson
Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington, DC 20001-1431
202-336-7075 phone
202-789-0596 fax
johno@ccsso.org

Burton Taylor
Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington, DC 20001-1431
202-336-7043 phone
202-371-1766 fax
burtont@ccsso.org

Phoebe Winter
Council of Chief State School Officers
2319 Traymore Road
Richmond, VA 23235
804-272-0996 phone
804-272-0677 fax
pwinter123@aol.com (preferred)